Representing CARE Rules in a Decision-theoretic Formalism

Michael M. Wagner,[1] J. Marc Overhage,[2] Eric Rodriguez,[3] and Gregory F. Cooper[1]
[1]Center for Biomedical Informatics, [3]Division of Geriatrics, University of Pittsburgh
[2]Regenstrief Institute and Department of Medicine, Indiana University School of Medicine

Improvement in the performance of reminder systems may be facilitated by the use of new representations. A decision-theoretic representation, for example, may enable a reminder system to represent and reason about the probabilities that a reminder will be a true or a false alarm and the relative utilities of these events. We extended a previously described decision-theoretic model to include such events. The model now represents explicitly the uncertainty, costs, and benefits of sending a reminder. We also extended the model to remove an assumption of reminder independence. As a step towards testing a hypothesis that this approach will support better performance than a rule-based approach, we analyzed a set of CARE rules and showed that our representation can represent these rules.

INTRODUCTION

After two decades of research, the basic results about the reminding paradigm—acceptability to clinicians, strong influence on clinician behavior, and ability to improve selected outcomes—are well established [1-9]. A key open problem at this juncture is how to improve the performance of reminder systems; that is, how to decrease the numbers of false alarms (which we define as reminders that recommend actions that are not appropriate for patients) while maintaining or increasing the numbers of true alarms.

The decision-theoretic formalism has potential for improving the performance of reminder systems. When used as the basis for medical expert systems, this formalism has produced improvements in performance relative to rule-based implementations [10]. Additionally, the decision-theoretic formalism can potentially model complex phenomena that affect reminder-system performance such as how clinicians respond when too many reminders are sent to them at once.

In our previous work, we formulated reminding as a decision problem [11], tested our approach in a laboratory setting [12], showed how to handle data error that is ubiquitous in field settings [13], and discussed how a decision-theoretic reminding knowledge base could address several problems related to sharing [13].

In this paper, we extend our previous model. We first develop a definition of reminder-system performance based on rates of true and false alarms and the utilities of these events. We then modify our original model—motivated by the objective of improving performance—to represent different types of true and false alarms and their utilities. The new model also relaxes an assumption of reminder-independence that was present in the previous model. Finally, we analyze a sample of CARE rules to demonstrate the representational adequacy of our model. The CARE language [14] has been used to generate reminders for over 20 years as part of the Regenstrief Medical Record System [15]. More recently, a version of the language called G-CARE is used to generated reminders in real time for a physician decision-support system [16]. CARE and G-CARE have been used for a large number of the high quality studies of reminder systems [17]; we assume that, at a minimum, a reminder-system language should have the expressiveness of these languages.

PERFORMANCE OF REMINDER SYSTEMS

Performance is an important issue for developers of reminder systems. Developers worry that too many false alarms will lead to clinician noncompliance or that too many true alarms might lead to unquestioning compliance with reminders [Reed Gardner, personal communication]. They have concerns that too many alarms of either type will overwhelm clinicians; therefore, sometimes they inactivate rules solely to decrease the volume of reminding. They employ rules-of-thumb such *as do not write a reminder rule that sends more than 66% false alarms* to convey their experience with the tradeoffs between true and false alarms to other rule authors.

Despite the importance of performance, current approaches to its measurement are not well-developed and have limitations that we will discuss. Moreover, there are no standard definitions of performance or methods for its measurement that we can apply to determine whether or not the performance of a reminder system is improved after a modification.

A typical approach to performance measurement is as follows: After a modification (e.g., an edit of a rule), we run the modified rule against a sample of historical patient cases and judge whether the

0195—4210/96/$5.00 © 1996 AMIA, Inc.

reminders being generated by the rule are improved (typically, the rule has been modified to exclude or include a target subset of patients). One limitation of this approach results from a fundamental tradeoff that exists between the rates of true and false alarms—that is, because the reminder system cannot precisely identify the target subset of patients, one rate can be improved only at the expense of the other. The current approach to performance evaluation assumes that developers are good judges of the effect of a change on the true- and false-alarm rates and on their relative utilities. A second limitation of current approaches is that the true value of a reminder system also depends on whether the reminders are heeded. For example, if reminder system A sends only true alarms, but none are heeded by clinicians and reminder system B sends the same true alarms but all are heeded, then a performance metric should rate B higher than A. The rates of unheeded reminders, incidentally, are not insignificant in working systems. From an analysis of published studies, we estimate that 30% of true alarms go unheeded [5, 7, 18]. A metric must also distinguish between false alarms that are heeded, and those that are not. False alarms that are heeded may be serious, as, for example, when a reminder is heeded to give penicillin to a patient who is allergic to penicillin, whereas false alarms that go unheeded usually just waste a clinician's time. Researchers have measured rates of unheeded false alarms of 19%, 22%, and 29% [5, 8, 18]. (We are not aware of published data about the rates at which false alarms are heeded.)

We have discussed these determinants of reminder-system performance because the decision-theoretic model discussed in the next section represents explicitly these types of alarms and their utilities. As an aside, we suggest that an ideal measure of reminder-system performance is a utility-weighted sum of these events.

A NEW DECISION-THEORETIC MODEL

In our previous model [12, 19], the utility of a reminder could take one of three values, depending on whether the reminder caused the clinician to do the wrong thing, nothing, or the right thing for a patient. In the new model, the utility of a reminder can take four possible values—corresponding to the events that the reminder is appropriate and heeded; appropriate and not heeded; inappropriate and heeded; or inappropriate and not heeded. This formulation has several advantages over the previous formulation. First, since performance is a function of those events, a normative system that reasons about them is guaranteed to exhibit optimal performance under well-defined assumptions. Second, methods for detecting and measuring these events have been developed by reminder-system researchers. Third, the utility of these events may be easier to estimate than the events used in our previous model.

Figure 1 is a decision tree that represents the computation performed by the new model for one reminder. In Figure 1, the decision alternatives are to send reminder r1 or to not send it. The outcome of sending r1 is that it is either appropriate and heeded, appropriate and not heeded, inappropriate and heeded, or inappropriate and not heeded. The reminder system's uncertainty about which it will be is represented by a probability distribution that is conditioned on the evidence available to the system, E. We model the utilities for these four outcomes as the differences between the utility of the action resulting from the reminder (e.g., the utility of doing the right thing or the wrong thing for a patient) and the cost of interrupting the clinician, denoted as *C(t)*, the utility of the time.

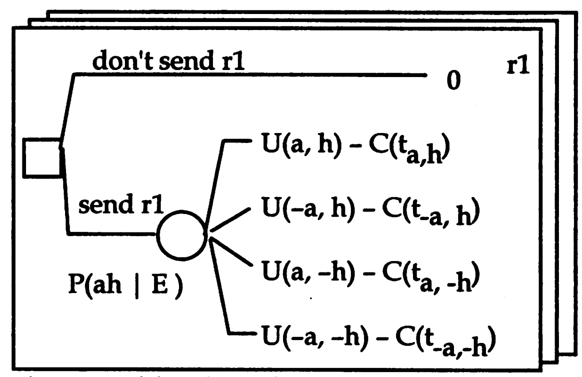

Fig. 1. Decision-theoretic model of reminding

Figure 2 shows a probabilistic model (a belief network) that can be used to compute the probability distribution P(ah | E). In this example, we use a hypothetical reminder that warfarin is appropriate therapy for chronic atrial fibrillation. This network models the uncertain relationships among what the data in an electronic medical record say about a patient *(warfarin, atrial fib)*, what the patient's true state is *(WARFARIN, ATRIAL FIB)*, whether the logic of a practice guideline about anticoagulation is satisfied *(logic satisfied)*, and whether a reminder will be appropriate and heeded by the clinician to whom it might be sent (ah). The local probability distributions for each node are shown adjacent to the node. Some features of this model are the explicit modeling of data accuracy, the encoding of the logic of a practice guideline within the probabilistic framework (i.e., as a subnetwork with a probability distribution that.

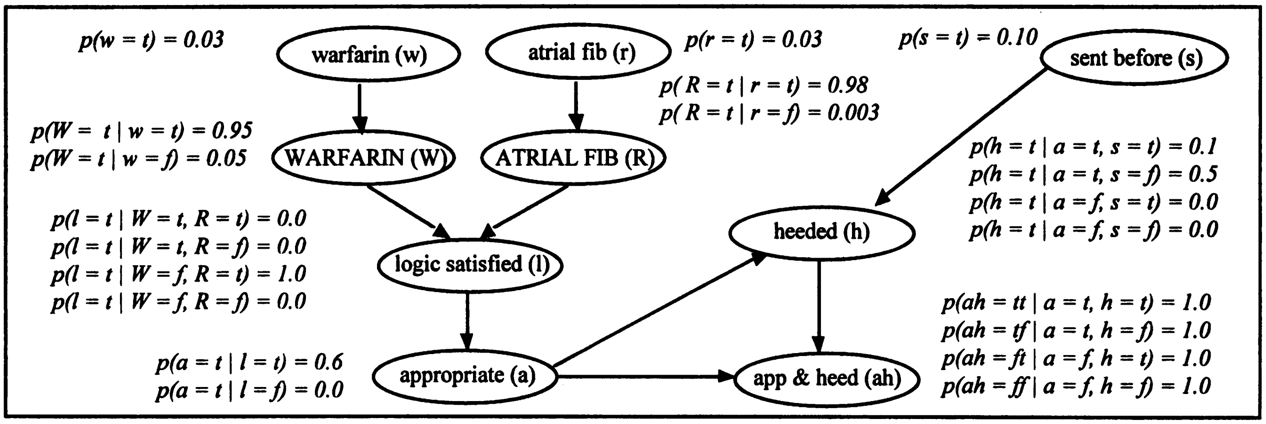

Fig. 2. A belief network that can compute the probability of the four outcomes of a reminder about warfarin therapy.

simulates the truth-value semantics of the logic underlying the practice guideline), the explicit modeling of practice guideline accuracy, and the explicit modeling of different types of true and false alarms

Our previous model assumed that the utility of a reminder is independent of other reminders sent. Based on an informal analysis of CARE rules, we modified the previous model in a second way. CARE rules often nest multiple reminders within a single logical structure to prevent the recommendation of two conflicting actions. We can model such dependencies by the addition of one subtree (analogous to the subtree for reminder 1 in Figure 1) for each nested reminder. When evaluated, the decision tree computes the expected utility of each of the decision alternatives (send no reminder, send r1, ... , send r_n) and acts on the alternative with the highest expected utility. The new model makes an assumption of modular independence: those reminders whose interactions are not explicitly represented are independent. We discuss a procedure for translating an arbitrary CARE rule to this formalism in the results section.

Decision theory offers many additional modeling possibilities which we outline in the discussion. The emphasis of the current model is on achieving performance improvement by explicit modeling of data error and by allowing a reminder system to decide whether to send a reminder based on a calculation of expected utility.

METHODS

The goal of this study was to determine whether the new model could represent all CARE rules.

To obtain a sample of CARE rules, we identified key problem areas in geriatrics by a consensus process involving five geriatricians who were not familiar with the CARE language. The topics identified included stroke prevention (e.g., reminders about blood pressure control and treatment of atrial fibrillation), depression, and drug related monitoring. One of the authors (JMO) identified rules in the Regenstrief Medical Record System that corresponded to these topics. We then analyzed the set of CARE rules to determine whether they could be represented in the decision-theoretic formalism.

RESULTS

JMO identified 21 CARE rules pertaining to 24 problem areas suggested by the clinicians. These rules comprised 184 distinct reminders (average 8.8 per rule, range 1 to 69). Analysis of this sample revealed the following basic rule structure:

Exclude A
Case statement
    Case B, then send reminder R1 and exit
    Case C, then send reminder R2 and exit
    Case D, then send reminder R3 and exit
    Case E, then call rule Z and exit

where upper-case letters A through E represent logical expressions (e.g., angina AND NOT hypertension) that may include atoms whose truth value is determined by other rules (backward chaining). The *exclude* statement, if satisfied, blocks the rule from evaluating. The case statement sends the first reminder whose precondition is satisfied. Thus, the behavior of a rule with this structure is to send at most one reminder (or take other action such as triggering the evaluation of another rule (a form of forward chaining).

A procedure to translate a CARE rule with this structure to our decision-theoretic model is as follows:

1. Create a single decision node.

2. Add a distinguished decision alternative *send no reminder* with utility = 0.
3. For each case in the case statement,
   - Add a decision alternative and a chance node as in Figure 1.
   - Condition the probability distribution for the chance node on the set of atoms in the case statements up to and including the current case (e.g., for R3, the elements of the expression NOT B AND NOT C AND D (i.e., create a belief network as in Figure 2 in which the logic of the case statement is duplicated).
   - For atoms that invoke backward chaining, define a variable that represents the result of the evaluation of that rule.
   - Add a utility function.
4. Check that the utility functions are consistent with the preferences implicit in the ordering.

The final step deserves additional explanation. In one CARE rule about the drug therapy of uncontrolled hypertension, all reminders to increase the dosage of ACE inhibitors appear in the case statement before any reminders to increase the dose of calcium-channel blockers, which in turn appear before the reminders for beta-blockers. This ordering within the case statement reflected a preference of the knowledge engineer for the use of these drugs (based on the clinical literature). In our decision-tree model, such preferences would be represented explicitly by a utility function that was consistent with them. An advantage of the decision-theoretic representation is that if our preferences for treatments changes, we do not have to manipulate the orderings in the case statement (we can express the new preferences directly).

Note that the model thus far described would exhibit behavior identical to that of the CARE rule. A potential advantage of the probabilistic version of the CARE rules is that, if we add data error and practice guideline error, the behavior of the reminder system may change (appropriately).

We did not encounter in our sample a second CARE structure in which all reminders from a set of *n* whose preconditions are satisfied are sent. We would map this structure into *n* decision trees.

DISCUSSION

A long-term objective of our research is to investigate the potential of the decision-theoretic formalism—which has been used extensively to develop medical expert systems [10, 21] but not to develop reminder systems—to improve the performance of reminder systems. In the absence of an existing common definition of performance, we developed one in this paper. We then developed a decision-theoretic model that will exhibit optimal performance (for a given set of reminders) as a consequence of computing with this formalism. We demonstrated that this formalism is at least as expressive as one of the rule languages in regular, ongoing use today. Since the Arden syntax [20] provides similar constructs (its design was influenced by CARE), there is every reason to believe that this conclusion also applies to Arden (note that procedures that compute temporal, or other predicates in CARE or in Arden can be used to determine the truth value of variables in this approach). In the future, we plan to investigate whether the ability of this language to represent uncertain medical knowledge will allow the encoding of practice guidelines that cannot be represented easily with current formalisms Although expressiveness of reminder-system languages is of current interest [22-25], decision-theoretic representations have received little attention.

We have not yet implemented the set of CARE rules that we analyzed; hence, future work will include the acquisition of the probabilities and utilities for this set of reminders. We expect that the probabilities required can be elicited from experts or estimated from data. In previous studies [12, 19], we were able to determine the utilities of reminders, hence we are optimistic that utilities can be specified. We note that performance equal to a rule-based reminder system can be achieved using default probabilities for data error that assume no data error, perfect practice guideline accuracy, and a utility function that is consistent with the ordering of reminders in the case statements.

Our future plans are to implement this set of reminders in a decision-theoretic reminder system, using belief-network technology, and to compare its performance to that of a comparable rule-based system. Such an experiment will test the effect of modeling data error and letting the system adjust the rate of true and false alarms. Langlotz [26] has pointed out that decision theory provides a rational basis for making such judgments; that is, decision theory allows us to break down the complex choice into simpler assessments such as how likely is it that the reminder will be appropriate for a patient with certain characteristics, what is the accuracy of the underlying data, and to what extent is the reminder likely to be ignored by patients or doctors. We conjecture that such questions are more natural for clinicians, and therefore can be answered with more accuracy by them.

Our future plans also include the investigation of the effect on performance of other techniques based on this formalism such as the explicit display of probabilities and expected utilities to users, the use of time-dependent utility functions to model urgent alerts, the ability to prioritize reminders for clinicians based on utility considerations, and ability to model patient preferences.

REFERENCES

1. Bradshaw KE. A computerized laboratory alerting system to warn of life threatening events [Ph.D. dissertation]: University of Utah,1988.
2. Fordham D, McPhee SJ, et al. The cancer prevention reminder system. MD Comput 1990;7:289-95.
3. Larsen RA, Evans RS, et al. Improved perioperative antibiotic use and reduced surgical wound infections through the use of computer decision analysis. Infect Cont Hosp Epidemiol 1989;10:316-20.
4. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. N Engl J Med 1976;295:1351-5.
5. McDonald CJ, Hui SL, et al. Reminders to physicians from an introspective computer medical record. Ann Intern Med 1984;100(1):130-8.
6. McDonald CJ. Use of a computer to detect and respond to clinical events: Its effect on clinician behavior. Ann Int Med 1976;84(2):162-7.
7. Ornstein SM, Garr DR, et al. Computer-generated physician and patient reminders. Tools to improve population adherence to selected preventive services. J Fam Pract 1991;32:82-90.
8. Rind DM, Safran C, et al. The effect of computer-based reminders on the management of hospitalized patients with worsening renal function. In: Clayton PD, ed. Proc. 15th SCAMC. Washington, DC: McGraw-Hill, 1993:28–32.
9. Safran C, Rind DM, et al. A clinical trial of a knowledge-based medical record. MEDINFO, 1995.
10. Heckerman D. Probabilistic Similarity Networks. Cambridge, MA: The MIT Press, 1990.
11. Wagner MM, Cooper GF. Decision-theoretic information pretrieval: A generalization of reminding. In: Safran C, ed. Proc. 17th SCAMC. Washington, DC: McGraw-Hill, 1993:512-16.
12. Wagner MM, Cooper GF. Evaluation of a belief-network–based reminder system that learns from utility feedback. JAMIA: Symposium Supplement. 1995:666-72.
13. Hogan WR, Wagner MM. Using belief-networks to enhance sharing of medical knowledge between sites with variations in data accuracy. JAMIA: Symposium Supplement. 1995:218-22.
14. McDonald CJ. Action-Oriented Decisions in Ambulatory Medicine. Chicago: Year Book Medical Publishers, Inc., 1981.
15. McDonald CJ, Tierney W, et al.. The Regenstrief Medical Record System: 20 years of experience in hospitals, clinics, and neighborhood health centers. MD Comput 1992;9:206-17.
16. Overhage J, Mamlin B, et al.. A tool for provider interaction during patient care: G-CARE. JAMIA: Symposium Supplement. 1995:178-82.
17. Johnston ME, Langston KB, et al. Effects of computer-based clinical decision support systems on clinician performance and patient outcome: A critical appraisal of research. Ann Int Med 1994;120:135-42.
18. Litzelman DK, Dittus RS, et al. Requiring physicians to respond to computerized reminders improves their compliance with preventive care protocols. J Gen Intern Med 1993;8:311-7.
19. Wagner MM. Decision-theoretic Reminder Systems that Learn from Feedback [Ph.D. dissertation]: University of Pittsburgh, 1995.
20. ASTM. E 1460 Standard specification for defining and sharing modular health knowledge bases (Arden syntax for medical logic systems). 1992 Annual Book of ASTM Standards. Philadelphia: American Society for Testing and Materials, 1992:539-87. vol 14.01).
21. Cooper GF. Current research directions in the development of expert systems based on belief networks. App Stoch Mod Data Anal 1989;5:39-52.
22. Shwe M, Sujansky W, Middleton B. Reuse of knowledge represented in the Arden syntax. In: Frisse M, ed. Proc 16th SCAMC. Baltimore, Maryland: McGraw-Hill, 1992:512-6.
23. Starret J, Xie G. Comparison of three knowledge representation formalisms for encoding the NCEP Cholesterol Guidelines. In: Ozbolt J, ed. Proc. 18th SCAMC. Washington, D.C.: Hanley and Belfus, Inc., 1994:792–6.
24. Starren JB, Hripcsak G, et al. Encoding a post-operative coronary artery bypass surgery care plan in the Arden syntax. Comput Biol Med 1994;24(5):411-6.
25. Das AK, Musen MA. A comparison of the temporal expressiveness of three database query methods. JAMIA: Symposium Supplement. 1995:331-7.
26. Langlotz CP, Shortliffe EH, Fagan LM. Using decision theory to justify heuristics. AAAI-86, 1986:215-19.